



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2019

Nonsmooth Kantorovich–Newton Methods: Hypotheses and Auxiliary Problems

Klatte, Diethard ; Kummer, Bernd

Abstract: In Newton's method $0 \in f(x_k) + G(x_k)(x_{k+1} - x_k)$ for solving a nonsmooth equation $f(x) = 0$, the type of approximation of f by some (generally multivalued) mapping G determines not only the convergence behavior of the method, but also the difficulty and the concrete form of the auxiliary problems. With $G(x) = \partial f(x)$ (Clarke's Jacobian) – like for locally convergent semismooth Newton methods – and for various other generalized "derivatives", the inclusion is a canonical one, i.e., it describes the usual Newton step if f is continuously differentiable near x_k . In our paper, we are interested in Kantorovich-type statements of convergence and study which meaningful hypotheses and auxiliary problems for particular pairs (f, G) may replace those of the classical smooth case. In particular, we point out – theoretically and by an example – why the related hypotheses cannot be checked for canonical methods even if f is piecewise linear.

DOI: <https://doi.org/10.1007/s10013-019-00348-4>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-182281>

Journal Article

Accepted Version

Originally published at:

Klatte, Diethard; Kummer, Bernd (2019). Nonsmooth Kantorovich–Newton Methods: Hypotheses and Auxiliary Problems. *Vietnam Journal of Mathematics*, 47(3):639-657.

DOI: <https://doi.org/10.1007/s10013-019-00348-4>

Nonsmooth Kantorovich-Newton Methods: Hypotheses and Auxiliary Problems

Diethard Klatte · Bernd Kummer

Received: 11 October 2018 / Revised: 24 January 2019 / Accepted: date

Abstract In Newton’s method $0 \in f(x_k) + G(x_k)(x_{k+1} - x_k)$ for solving a nonsmooth equation $f(x) = 0$, the type of approximation of f by some (generally multivalued) mapping G determines not only the convergence behavior of the method, but also the difficulty and the concrete form of the auxiliary problems. With $G(x) = \partial f(x)$ (Clarke’s Jacobian) – like for locally convergent semismooth Newton methods – and for various other generalized “derivatives”, the inclusion is a *canonical* one, i.e., it describes the usual Newton step if f is continuously differentiable near x_k . In our paper, we are interested in Kantorovich–type statements of convergence and study which meaningful hypotheses and auxiliary problems for particular pairs (f, G) may replace those of the classical smooth case. In particular, we point out – theoretically and by an example – why the related hypotheses cannot be checked for canonical methods even if f is piecewise linear.

Keywords Nonsmooth Newton method · Kantorovich–type convergence · nonlinear approximation · auxiliary problem · convergence conditions · piecewise linear systems

Mathematics Subject Classification (2010) 49J53 · 49K40 · 90C31 · 65J05

1 Introduction

Kantorovich-type statements for Newton’s method are devoted in [22] to the standard approach for solving the equation $f = 0$ with a C^1 –function $f : X \rightarrow Y$ (X, Y Banach spaces) via the iterations (for $k = 0, 1, \dots$)

$$\Sigma(x_{k+1}, x_k) := f(x_k) + Df(x_k)(x_{k+1} - x_k) = 0, \quad x_0 \text{ given.} \quad (1.1)$$

The “usual”, *local convergence* theorem says that, for all initial points x_0 near a solution \bar{x} with existing $Df(\bar{x})^{-1}$, the generated sequence of possible auxiliary solutions to (1.1)

$$x_{k+1} \in \mathcal{S}(x_k), \quad x_0 \text{ given,} \quad (1.2)$$

is uniquely defined and converges superlinearly to \bar{x} . *Kantorovich-type convergence* (also called *semilocal convergence*) means that the hypothesis of small $\|x_0 - \bar{x}\|$ is replaced by the observable

Diethard Klatte
Institut für Betriebswirtschaftslehre, Universität Zürich, Moussonstrasse 15, CH-8044 Zürich, Switzerland. E-mail: diethard.klatte@uzh.ch

Bernd Kummer
Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany. E-mail: kummer@math.hu-berlin.de

hypothesis of small $\|f(x_0)\|$, and (linear or better) convergence to some solution $x^* \in f^{-1}(0) \cap \Omega$ is asserted provided that

$$\begin{aligned} & \text{the variation of the (sufficiently regular) derivatives} \\ & \text{on some (sufficiently large) convex region } \Omega \ni x_0 \text{ is small enough.} \end{aligned} \quad (1.3)$$

Hence the *method* remains the same, but the *assertions* change since, a-priori, even the existence of a solution in Ω is not supposed. Having $f \equiv f(x_0) \neq 0$ or the real function $f(x) = e^x$ in mind, assumptions concerning f^{-1} and the derivatives on Ω are obviously needed.

As usually, for extensions from smooth to nonsmooth models, the notion of derivatives has been generalized in the literature, and (1.1)–(1.3) must be adapted and concretized.

The nonsmooth model. For possibly nonsmooth functions $f : X \rightarrow Y$, we shall study any method which solves (for $k = 0, 1, \dots$)

$$0 \in \Sigma(x_{k+1}, x_k), \quad x_0 \text{ given}, \quad (1.4)$$

$$\text{where } \Sigma : X \times X \rightrightarrows Y \text{ is a multifunction with} \quad (1.5)$$

$$\Sigma(x, x) = \{f(x)\}. \quad (1.6)$$

Note that (1.1) fits into (1.4)–(1.6). This model is general enough to describe (not only) generalized Newton methods and so-called inexact Newton methods as well. The set $\mathcal{S}(x)$ now denotes the solutions of $0 \in \Sigma(\cdot, x)$ throughout the paper. Often one defines

$$\begin{aligned} \Sigma(x_{k+1}, x_k) &= f(x_k) + G(x_k)(x_{k+1} - x_k) \\ \text{and solves accordingly } 0 &\in f(x_k) + G(x_k)(x_{k+1} - x_k) \end{aligned} \quad (1.7)$$

where $G(x) : X \rightrightarrows Y$ is some "generalized derivative" of f at x (contingent derivative, directional derivative or something else) and $G(x)(u)$ consists of its values (approximated or not) for direction $u \in X$. By our definition, $G(x)$ need not to be positively homogenous.

A Kantorovich type theorem presents constants $r > 0$ and $\alpha > 0$ such that the method with start at x_0 finds a zero of f in $\Omega = B(x_0, r)$ if $\|f(x_0)\| < \alpha$. Notice that this is an implicit function theorem for f if only the existence of some $x^* \in f^{-1}(0) \cap B(x_0, r)$ is asserted and procedures, to find it, play no role.

If we fix $G(x) = G(x_0)$ for all x , (1.7) becomes a so-called *modified Newton method* which is closely related to the general procedure. In any case, assumption (1.6) means $G(x)(0) = \{0\}$, and the inverse generalized derivative $G(x)^{-1}$ assigns, to $v \in Y$, all directions $u \in X$ with $v \in G(x)(u)$; usually a nontrivial multifunction.

Obviously, the type of approximation of f by G determines not only the behavior of the method (theoretically and numerically) but also the difficulty and the content of the auxiliary problems. However, with the "best approximation" $G(x)(x' - x) := f(x') - f(x)$, which yields $\Sigma(x_{k+1}, x_k) = \{f(x_{k+1})\}$ in (1.7) and corresponds to

$$G(x)(u) = D^0 f(x)(u) := f(x + u) - f(x) \quad (\text{no derivative at all}), \quad (1.8)$$

the method (1.7) converges, but is useless since it requires to solve $f(x_{k+1}) = 0$.

The appropriate definitions of G for given f and their *applicability* for deriving Kantorovich-type statements form the subject of our paper. Thus we study which meaningful hypotheses and auxiliary problems for particular pairs (f, G) may replace those of the classical approach, when nonsmooth models are considered. In particular, we discuss potentials and limitations of results which are known for the studied subject. Our message is that, in this context, only few definitions of $G(x)$ (i.e., only few generalized methods) seem to be useful for *solving* the equation via (1.4) or (1.7) by a concrete method. This may be surprising since, influenced by the perturbation theory of variational analysis, there exist meanwhile many papers which present Kantorovich-type statements to seemingly various types of nonsmooth functions and generalized equation as well, [1, 5, 8, 9, 24, 31, 35, 37].

Like Robinson's concept of point-based approximations (PBAs) [35, 37] (1988, 1994) they are usually based on more or less abstract conditions for f and G . However, up to the few known

realizations for (f, G) where the auxiliary systems are at least piecewise linear, the content of the hypotheses and meaning of the auxiliary problems have not been discussed so far. So it is not clear which other pairs (f, G) satisfy the imposed conditions and what remains to do for solving the auxiliary systems (1.4) or (1.7). It is even open whether *canonical* methods, i.e., such ones which use the standard linearization if f is continuously differentiable near some iteration point (see Assumption 1 in §3.3), can be applied in that context. Our contribution will give a negative answer when (1.7) describes a general PBA and also when it stands for more general concepts we found, and this for both Newton and modified Newton methods as well.

The paper is organized as follows. In section 2, we present our notation, handle specifications of the model (1.4)–(1.6) and introduce some Lipschitz stability concepts which are used in our paper. After summarizing some standard auxiliary results for convergence, we study in section 3 the basic tools for local and semi-local convergence of generalized Newton methods under the theoretical point of view. In particular, we point out the role of the mean-value theorem and its generalizations, discuss the needed properties for approximations of the problem function f and deal with Robinson's [37] concept of PBAs and its extensions. As a consequence, sufficient convergence conditions under metric regularity are investigated. Section 4 is devoted to necessary convergence conditions and will show that the tools elaborated in the previous section can be only applied to few (mostly already well-known) nonsmooth models.

2 Preliminaries

Notation

The following notation is employed. Given Banach spaces X, Y , a mapping (multifunction) $\Gamma : X \rightrightarrows Y$ assigns, to each $x \in X$, some subset $\Gamma(x) \subset Y$ which may be empty. The inverse is given by $\Gamma^{-1}(y) = \{x \mid y \in \Gamma(x)\}$, and $\text{gph } \Gamma = \{(x, y) \mid y \in \Gamma(x)\}$ is the graph of Γ . The mapping Γ is called closed at $(x, y) \in X \times Y$ if the convergences $x' \rightarrow x$, $y' \rightarrow y$ for $y' \in \Gamma(x')$ imply $y \in \Gamma(x)$. In particular, this happens if Γ is closed which means that $\text{gph } \Gamma$ is closed in $X \times Y$. If nothing else is said we will throughout assume that elements x, u belong to X while y, v are in Y .

By $\mathcal{L}(X, Y)$ we denote the space of linear, continuous functions $A : X \rightarrow Y$ with its usual norm. $B_X(x, r)$ is the closed ball with radius r around x , similarly $B_X(M, r)$ is the union of all $B_X(x, r)$ for $x \in M$. Mostly the space is evident, then we omit the index.

The sum $M_1 + M_2$ of sets $M_1, M_2 \subset X$ denotes Minkowski's sum $M = \{m_1 + m_2 \mid m_1 \in M_1, m_2 \in M_2\}$. If $M_1 = \{m_1\}$, we also write $M = m_1 + M_2$. The Hausdorff-distance of M_1 and M_2 is defined by $d_H(M_1, M_2) = \inf\{\varepsilon > 0 \mid M_1 \subset B(M_2, \varepsilon), M_2 \subset B(M_1, \varepsilon)\}$.

We say a statement holds near x if it holds for all x' in some neighborhood of x . For $f : X \rightarrow Y$ and $\emptyset \neq \Omega \subset X$, we denote by $\text{Lip}(f, \Omega)$ the smallest Lipschitz modulus for f on Ω , including ∞ if f is not Lipschitz. Similarly, $\sup(f, \Omega) = \sup_{x \in \Omega} \|f(x)\|$. We call x a C^1 -point of f if the (Fréchet) derivative Df exists and is continuous near x .

Model settings

Remark 1 We first note that supposing (1.7) is no restriction of generality when studying the model (1.4)–(1.6). The representation (1.7) is possible for arbitrary Σ since $\Sigma_f(x', x) := \Sigma(x', x) - f(x) \subset Y$ depends uniquely on x and $x' - x$ and allows to define $G(x)(x' - x) := \Sigma_f(x', x)$.

The following examples demonstrate particular settings in (1.4)–(1.6).

1. To apply generalized Jacobians in \mathbb{R}^n , [6, 31], put $G(x)(u) = \{Au \mid A \in \partial f(x)\}$.
2. To apply "Newton maps" [23, 24], replace $\partial f(x)$ by appropriate nonempty subsets $\mathcal{N}f(x)$ of the space of linear, continuous functions from X to Y .
3. Let $Y = X$ and define, for any given $S : X \rightrightarrows X$, the mapping $\Sigma : X \times X \rightrightarrows X$ by $\Sigma(x', x) = x' - S(x)$. Then $x' \in S(x)$ and $0 \in \Sigma(x', x)$ are equivalent and condition $\Sigma(x, x) = \{f(x)\}$ becomes $x - S(x) = \{f(x)\}$. Hence $S(x)$ is automatically single-valued and $x' \in S(x)$ coincides with

$x' - x = -f(x)$. Thus $\Sigma_f(x', x) := \Sigma(x', x) - f(x) = x' - S(x) - f(x) = x' - x$, $G(x)$ is the identity and method (1.7) turns into $x_{k+1} = x_k - f(x_k)$. Having this already for smaller k , one obtains $x_{k+1} - x_0 = (x_{k+1} - x_k) + (x_k - x_{k-1}) + \dots + (x_1 - x_0) = -[f(x_0) + \dots + f(x_k)]$, and convergence of the iterates means exactly convergence of the sum of function values.

Generally, one easily observes for the solutions of the general model (1.4)–(1.7):

Remark 2 The iterates x_{k+1} in (1.7) are explicitly given by

$$x_{k+1} \in \mathcal{S}(x_k) := \Sigma(\cdot, x_k)^{-1}(0) = x_k + G(x_k)^{-1}(-f(x_k)). \quad (2.1)$$

Thus the solutions $x \in f^{-1}(0)$ are just the fixed points of the composed mapping

$$x \mapsto \mathcal{S}(x) = x + G(x)^{-1}(-f(x)) \quad (2.2)$$

due to $G(x)(0) = \{0\}$ and $x \in \mathcal{S}(x) \Leftrightarrow 0 \in G(x)^{-1}(-f(x)) \Leftrightarrow -f(x) \in G(x)(0) \Leftrightarrow f(x) = 0$.

Lipschitz properties of multifunctions

Our terminology for local Lipschitz and regularity concepts follows that of [23]. We refer e.g., to [4, 6, 7, 10, 13, 17, 18, 23, 27, 28, 34, 38, 39] for the rich literature on these concepts, including equivalent definitions, characterizations, covering properties and their interrelations.

Definition 1 A multifunction $F : Y \rightrightarrows X$ is *pseudo-Lipschitz*, in short (*psLip*), with rank $\kappa > 0$ at $(\bar{y}, \bar{x}) \in \text{gph } F$ and $\Gamma = F^{-1}$ is called *metrically regular* or *pseudo-regular* at (\bar{x}, \bar{y}) if there are positive r_1, r_2 such that

$$\begin{aligned} &\text{there is some } x' \in F(y') \cap B_X(x, \kappa \|y' - y\|) \\ &\text{if } x \in F(y) \cap B_X(\bar{x}, r_1) \text{ and } y, y' \in B_Y(\bar{y}, r_2). \end{aligned} \quad (2.3)$$

Often, instead of the name “pseudo-Lipschitz” which goes back to [4], the notions *Aubin property* [38] or *Lipschitz like* [28] are used. The mapping $\Gamma = F^{-1}$ is said to be *strongly regular* (or *strongly metrically regular* [10]) at (\bar{x}, \bar{y}) if r_1, r_2 and κ exist in such a way that (2.3) holds and x' is unique for the corresponding points. Then the restriction \hat{F} of F to the balls is a *function* with $\text{Lip}(\hat{F}, B_Y(\bar{y}, r_2)) \leq \kappa$, i.e., F has a *Lipschitz continuous single-valued localization around \bar{y}* in the sense of [10].

Definition 2 A multifunction $F : Y \rightrightarrows X$ is *locally upper Lipschitz*, in short (*luLip*), with rank $\kappa > 0$ at $(\bar{y}, \bar{x}) \in \text{gph } F$ if there are positive r_1, r_2 such that

$$F(y) \cap B_X(\bar{x}, r_1) \subset B_X(\bar{x}, \kappa \|y - \bar{y}\|) \text{ holds for all } y \in B_Y(\bar{y}, r_2). \quad (2.4)$$

Instead of (luLip) also the notion *isolated calmness* is used in the literature, and *strong metric subregularity* of $\Gamma = F^{-1}$ stands for F being (luLip).

For several estimates, the size of the “*stability balls*” $B_X(\bar{x}, r_1), B_Y(\bar{y}, r_2)$ plays a role.

Notes.

(N1) The radii r_1, r_2 restrict the region in $X \times Y$ where the definition may be used.

(N2) Metric and strong regularity are *persistent* with respect to small variations of $(\bar{x}, \bar{y}) \in \text{gph } \Gamma$.

(N3) If F is a *function* with $\text{Lip}(F, B_Y(\bar{y}, r_2)) \leq \kappa$ then (2.3) and (2.4) hold with any $r_1 > 0$.

(N4) Condition (2.3) is valid if $d_H(F(y'), F(y)) \leq \kappa \|y' - y\| \forall y, y' \in B_Y(\bar{y}, r_2)$. Condition (2.4) then holds if $F(\bar{y}) = \{\bar{x}\}$.

(N5) A linear operator $F \in \mathcal{L}(Y, X)$ is both (psLip) and (luLip) at all $(\bar{y}, \bar{x}) \in \text{gph } F$ with $\kappa = \|F\|$ and arbitrary r_1, r_2 . Any $\Gamma \in \mathcal{L}(X, Y)$ is metrically regular if $\Gamma(X) = Y$ and strongly regular if $\Gamma^{-1} \in \mathcal{L}(Y, X)$ exists.

3 Basic tools and convergence conditions

In this section, the basic tools for deriving convergence conditions in generalized Newton methods are discussed. To point out both potentials and limitations in the nonsmooth case, this is brought into relation to classical ideas used in the smooth setting.

3.1 Auxiliary results

The first two lemmas are standard in the study of convergence of procedures as considered in our paper. Though known from the related literature (cf. e.g. [1, 3, 22, 29, 32]), we present their proofs for the reader's convenience.

In order to ensure convergence at all by the contraction principle one can easily show without using additional assumptions like (1.4)–(1.7):

Lemma 1 (i) *Let $S : X \rightrightarrows X$ be arbitrary and let $\{x_k\}$ be any sequence satisfying, for given x_0 and $q \in (0, 1)$,*

$$x_{k+1} \in S(x_k) \quad \text{and} \quad \|x_{k+1} - x_k\| \leq q \|x_k - x_{k-1}\| \quad (k = 1, 2, \dots). \quad (3.1)$$

Then the sequence converges in the ball $\Omega_q := B(x_0, \frac{\|x_1 - x_0\|}{1-q})$ to some x^ .*

(ii) *If $q \in (0, \frac{1}{3})$, the convergence is linear, i.e.,*

$$\|x_{k+1} - x^*\| \leq \theta \|x_k - x^*\| \quad \text{holds for some } \theta < 1 \quad \text{and} \quad k = 0, 1, 2, \dots \quad (3.2)$$

(iii) *A sequence satisfying (3.1) for $q \in (0, 1)$ exists if*

$$\begin{aligned} &\text{for some } x_1 \in S(x_0) \text{ and all } x, x' \in \Omega_q \text{ there holds} \\ &\emptyset \neq S(x) \cap B(x, q \|x - x'\|) \text{ provided that } x \in S(x'). \end{aligned} \quad (3.3)$$

Proof (i) By (3.1) one obtains the estimate of Banach's fixed point theorem

$$\|x_0 - x_{k+1}\| \leq \|x_0 - x_1\| + \dots + \|x_k - x_{k+1}\| \leq (1 + q + q^2 + \dots) \|x_1 - x_0\| = \frac{\|x_1 - x_0\|}{1 - q}. \quad (3.4)$$

Hence all x_k stay in Ω_q and form a geometrically convergent Cauchy sequence by boundedness of the sum. In consequence, $\lim x_k = x^*$ exist.

(ii) Linear convergence for $q \in (0, \frac{1}{3})$ has been shown in [1, p.181]. We add a proof since there are two inessential but disturbing index-errors in the first line. By (3.4) one obtains

$$\begin{aligned} \|x_k - x^*\| &= \lim_{m \rightarrow \infty} \|x_k - x_m\| \\ &\leq \lim_{m \rightarrow \infty} (\|x_k - x_{k+1}\| + \|x_{k+1} - x_{k+2}\| + \dots + \|x_{k+m-1} - x_m\|) \\ &\leq \|x_k - x_{k+1}\| (1 + q + q^2 + \dots) \\ &\leq \frac{\|x_k - x_{k+1}\|}{1 - q} \leq \frac{q}{1 - q} \|x_{k-1} - x_k\| \leq \frac{q}{1 - q} (\|x_{k-1} - x^*\| + \|x^* - x_k\|). \end{aligned}$$

Thus

$$(1 - \frac{q}{1 - q}) \|x_k - x^*\| = \frac{1 - 2q}{1 - q} \|x_k - x^*\| \leq \frac{q}{1 - q} \|x_{k-1} - x^*\|$$

presents (3.2) with $\theta = q(1 - 2q)^{-1} < 1$.

(iii) To verify (iii) for given $x_1 \in S(x_0)$, it suffices to put $x = x_k$ and $x' = x_{k-1}$ in (3.3) for $k \geq 1$. This yields $\emptyset \neq S(x_k) \cap B(x_k, q \|x_k - x_{k-1}\|)$. Hence some $x_{k+1} \in S(x_k)$ satisfying (3.1) exists. Since $x_{k+1} \in \Omega_q$ follows again from (3.4), the proof is complete.

Another classical way to verify convergence uses majorizing sequences.

Lemma 2 *Given $S : X \rightrightarrows X$ let $\{x_k\}$ be any sequence satisfying, for given x_0 ,*

$$x_{k+1} \in S(x_k) \quad \text{and} \quad \|x_{k+1} - x_k\| \leq t_{k+1} - t_k \quad (k = 0, 1, 2, \dots) \quad (3.5)$$

where $t_k \rightarrow t^$ is a real monotone sequence with $t_0 = 0$. Then the sequence $\{x_k\}$ converges in the ball $B(x_0, t^*)$ to some x^* .*

Proof Now one obtains a Cauchy sequence in $B(x_0, t^*)$ due to $\|x_{k+1} - x_0\| \leq \|x_{k+1} - x_k\| + \dots + \|x_1 - x_0\| \leq (t_{k+1} - t_k) + \dots + (t_1 - t_0) = t_{k+1} - t_0 \leq t^*$.

Remark 3 In both Lemmas, it follows $x^* \in S(x^*)$ if S is closed at (x^*, x^*) .

It is important how (and whether) the hypotheses of these Lemmas can be verified. The uniqueness of $x_{k+1} \in S(x_k)$ will be an extra gift under many settings (using strong regularity).

The following version of Banach's Perturbation Lemma (cf. [22] p. 140) is a basic auxiliary result for applying the Lemmas 1 and 2 to C^1 - equations and method (1.1), (1.2).

Lemma 3 *Let $f \in C^1(X, Y)$ and $\Omega \ni x_0$ be convex. Suppose that $U^{-1} := Df(x_0)^{-1}$ exists*

and positive L, δ with $L\delta < 1$ satisfy $\|U^{-1}\| \leq L$ and $\sup_{x \in \Omega} \|Df(x) - Df(x_0)\| \leq \delta$. (3.6)

Then $Df(x)^{-1}$ exists for all $x \in \Omega$ and fulfills

$$\|Df(x)^{-1}\| \leq \alpha_1(x) := \frac{\|U^{-1}\|}{1 - \|U^{-1}(Df(x) - Df(x_0))\|} \quad (3.7)$$

$$\leq \alpha_2(x) := \frac{L}{1 - L\|Df(x) - Df(x_0)\|} \leq \Lambda_\delta := \frac{L}{1 - L\delta}. \quad (3.8)$$

Note that $\|U\| = \text{Lip}(U, X)$ is the smallest (psLip) and (luLip) rank of U at $(0, 0)$. So the Lemma says that $Df(x)^{-1}$ is both (psLip) and (luLip) with rank Λ_δ at the origin. Lemma 3 ensures various statements on Newton's method including Prop. 1 in Section 3.2.

3.2 The mean-value theorem and needed properties of approximations

For deriving *necessary and/or sufficient convergence-conditions*, the *mean value theorem* plays a crucial role. In an elementary manner, this can be seen as follows. Knowing that (1.7) and (2.1) hold already for smaller k , the iterates fulfill $0 \in f(x_{k-1}) + G(x_{k-1})(x_k - x_{k-1})$, i.e.,

$$-f(x_k) \in -f(x_k) + f(x_{k-1}) + G(x_{k-1})(x_k - x_{k-1}) \quad (k \geq 1). \quad (3.9)$$

Whenever x_{k+1} for $k \geq 1$ exists then, substituting for $-f(x_k)$ in (2.1), we obtain

$$x_{k+1} - x_k \in \Psi_k := G(x_k)^{-1} [f(x_{k-1}) - f(x_k) + G(x_{k-1})(x_k - x_{k-1})] \quad (k \geq 1). \quad (3.10)$$

To ensure convergence of iterates x_k in any set Ω at all, the norms $\|x_{k+1} - x_k\|$ must vanish. Therefore, the pair (f, G) and the elements $\psi_k := x_{k+1} - x_k$ have to fulfill

$$\exists v_k \in f(x_{k-1}) - f(x_k) + G(x_{k-1})(x_k - x_{k-1}) \quad \text{and} \quad \psi_k \in G(x_k)^{-1}(v_k) \quad \text{such that} \quad \psi_k \rightarrow 0. \quad (3.11)$$

In order to speak about a "method", we must be also able to identify these elements.

The smooth case

For $f \in C^1$ the classical method uses $G(x) = Df(x)$ and $G(x)^{-1} = Df(x)^{-1}$, provided the linear inverse operator exists, and (3.10) is the well-known formula

$$x_{k+1} - x_k = Df(x_k)^{-1} [f(x_{k-1}) - f(x_k) + Df(x_{k-1})(x_k - x_{k-1})] \quad (k \geq 1). \quad (3.12)$$

Here, v_k and ψ_k are unique and the mean value theorem guarantees that

$$\|v_k\| = o(\|x_k - x_{k-1}\|) \quad \text{holds for converging} \quad x_k \rightarrow x^*. \quad (3.13)$$

Clearly, if $\|Df(x)^{-1}\|$ is bounded on Ω , say

$$\|Df(x)^{-1}\| \leq \Lambda \quad \text{for all } x \in \Omega \quad (3.14)$$

as under Lemma 3, one obtains, from $\|\psi_k\| \leq \Lambda o(\|x_k - x_{k-1}\|) \leq q \|x_k - x_{k-1}\|$, already

$$\|x_{k+1} - x_k\| \leq q \|x_k - x_{k-1}\| \text{ with some } q \in (0, 1), \quad (3.15)$$

at least for small $\|x_k - x^*\|$. These facts are the key for showing *local superlinear convergence* of the classical Newton method.

To obtain a *Kantorovich-type theorem*, $\|v_k\|$ can be globally estimated, for points in Ω , via (Banach's Perturbation) Lemma 3 which connects $\delta := \sup_{x \in \Omega} \|Df(x) - Df(x_0)\|$, $\|Df(x_0)^{-1}\|$ and Λ . In particular, one has

$$\text{if } \|Df(x_0)^{-1}\| \leq L \text{ and } L\delta < 1 \text{ then (3.14) holds with } \Lambda \leq \frac{L}{1-L\delta}.$$

So Lemma 1 and 3 are also the basic tools for proving the subsequent result on (geometric) convergence, which is more or less explicitly contained in all papers on Kantorovich theorems for C^1 functions as, e.g., in [22] and [29].

Proposition 1 *Let $f \in C^1(X, Y)$ and $\Omega \ni x_0$ be convex. Suppose that $U^{-1} := Df(x_0)^{-1}$ exists and positive L, δ with $L\delta < 1$ satisfy along with some $x_1 \in \mathcal{S}(x_0)$*

$$\|U^{-1}\| \leq L, \quad \|Df(x') - Df(x)\| \leq \delta \quad \forall x', x \in \Omega, \quad (3.16)$$

$$q := \frac{L\delta}{1-L\delta} < 1 \quad \text{and} \quad \Omega_q := B(x_0, \frac{\|x_1 - x_0\|}{1-q}) \subset \Omega. \quad (3.17)$$

Then sequence (1.1) is well-defined and converges in Ω geometrically to some zero x^ of f .*

Notice that $\Lambda_\delta = q/\delta$ in Lemma 3, and $q < 1 \Leftrightarrow L\delta < \frac{1}{2}$.

Proof The point $x_1 \in \mathcal{S}(x_0)$ belongs trivially to $\Omega_q \subset \Omega$. Hence $Df(x_1)^{-1}$ exists by Lemma 3. Beginning with $k = 1$, the existence of $Df(x_k)^{-1}$ implies that $x_{k+1} \in \mathcal{S}(x_k)$ uniquely exists, namely $x_{k+1} = x_k - Df(x_k)^{-1}f(x_k)$. To exploit Lemma 1 one can verify (3.1) for $S = \mathcal{S}$ by applying the *mean value theorem*. Indeed, the estimates $\alpha_\nu(x_k)$ of $\|Df(x_k)^{-1}\|$ in Lemma 3 and (3.12) ensure

$$\|x_{k+1} - x_k\| = \|Df(x_k)^{-1} [f(x_k) - f(x_{k-1}) - Df(x_{k-1})(x_k - x_{k-1})]\| \quad (3.18)$$

$$\leq \alpha_\nu(x_k) \left\| \int_0^1 Df(x_{k-1} + \tau(x_k - x_{k-1}))(x_k - x_{k-1}) d\tau - Df(x_{k-1})(x_k - x_{k-1}) \right\| \quad (3.19)$$

$$\leq \alpha_\nu(x_k) \left(\sup_{\tau \in (0,1)} \|Df(x_{k-1} + \tau(x_k - x_{k-1})) - Df(x_{k-1})\| \right) \|x_k - x_{k-1}\| \quad (3.20)$$

$$\leq \Lambda_\delta \delta \|x_k - x_{k-1}\| = q \|x_k - x_{k-1}\|. \quad (3.21)$$

Thus $Df(x_{k+1})$ exists again by (3.4) and $x^* = \lim x_k \in \Omega$ exists by Lemma 1. Since $f \in C^1$ and $Df(x_k) \rightarrow Df(x^*)$, it follows closedness of \mathcal{S} at (x^*, x^*) and $f(x^*) = 0$ as well.

Notes.

1. If $q < 1/3$, one obtains also linear convergence to x^* , see Lemma 1.
2. Since $\|x_1 - x_0\| = \|Df(x_0)^{-1}f(x_0)\| \leq L\|f(x_0)\|$, small values $\|f(x_0)\|$, such that $B(x_0, \frac{L\|f(x_0)\|}{1-q}) \subset \Omega$, are in fact sufficient in order to obtain $\Omega_q \subset \Omega$ and the claimed convergence for given $q < 1$.
3. If $\bar{x} \in f^{-1}(0)$ and $Df(\bar{x})^{-1}$ exists, all hypotheses of Prop. 1 can be satisfied for sufficiently small $\|x_0 - \bar{x}\|$. Therefore, Kantorovich's approach generalizes local statements on the convergence of Newton's method.
4. If $x \mapsto Df(x)$ is Lipschitz with rank β on $\Omega = B(x_0, \frac{\|x_1 - x_0\|}{1-q})$, one may put $\delta = 2\beta \frac{\|x_1 - x_0\|}{1-q}$ in (3.16), after which (3.17) holds for small β , too: require $2L\beta \frac{\|x_1 - x_0\|}{1-q} < \frac{1}{2}$. For $f \in C^2$, it holds $\beta \leq \sup_{x \in \Omega_q} \|D^2f(x)\|$. Though not sharp, it is hard to find a better estimate.

The proof of Prop. 1 permits several variations if $x \mapsto Df(x)$ is Lipschitz as above. By

$$\|Df(x_{k-1} + \tau(x_k - x_{k-1})) - Df(x_{k-1})\| \leq \beta \|x_k - x_{k-1}\| \quad (3.22)$$

in (3.19), one obtains with $\alpha_2(x_k)$ the quadratic estimate

$$\|x_{k+1} - x_k\| \leq \frac{L \left(\frac{1}{2} \beta \|x_k - x_{k-1}\|^2 \right)}{1 - L \|Df(x_k) - Df(x_0)\|} \leq \frac{1}{2} \frac{L\beta}{1 - L\beta \|x_k - x_0\|} \|x_k - x_{k-1}\|^2$$

Without going into the details, now the hypothesis

$$\lambda := L\beta \|x_1 - x_0\| < \frac{1}{2} \quad (3.23)$$

becomes the crucial condition for complete induction in [29]. The latter is also the key for estimates in [22] and other papers which state quadratic convergence via majorizing sequences according to Lemma 2. There, due to the Lipschitz property (3.22) of Df , the radius t^* and possible t_k are indeed *available*: One may choose t_k as the k -th Newton iterate for the real function $\phi(t) = \frac{1}{2}L\beta t^2 - t + \|x_1 - x_0\|$. This implies that $t^* = (\beta L)^{-1}(1 - \sqrt{1 - 2\lambda})$ is the smallest zero of ϕ . We do not follow this line since we are interested in nonsmooth f .

The nonsmooth case

For $f \notin C^1$ in our general model (1.4)–(1.7), mean value theorems are weaker or do not hold at all, and $G(x), G(x)^{-1}$ in (3.10) are no more single-valued in general. The crucial assumption (3.14) now corresponds to

$$G(x)^{-1}(v) \neq \emptyset \quad \text{and} \quad \|u\| \leq A \|v\| \quad \text{for all } x \in \Omega \quad \text{and} \quad u \in G(x)^{-1}(v). \quad (3.24)$$

In a weaker local form, for any $r_2 > 0$, $r_1 = \Lambda r_2$ and $\|v\| < r_2$, this means in terms of Lipschitz properties of multifunctions in §2 that

$$\emptyset \neq G(x)^{-1}(v) \subset B(0, \Lambda \|v\|) \subset B(0, r_1) \quad \forall v \in B(0, r_2), \quad x \in \Omega,$$

hence,

$$G(x)^{-1}(v) \cap B(0, r_1) \neq \emptyset \quad \text{and} \quad G(x)^{-1} : Y \rightrightarrows X \text{ is (luLip) with rank } \Lambda \text{ at } (0, 0), \text{ uniformly for all } x \in \Omega. \quad (3.25)$$

Having the mean-value statement (3.13) along with (3.24) and non-empty images of $G(x)^{-1}$, then one obtains the contractivity (3.15) again as above at least for small $\|x_k - x^*\|$.

Remark 4 These facts are the key for showing *local convergence* of nonsmooth Newton methods. They essentially apply that (3.13) holds for $G(x) = \partial f(x)$ if f is semismooth in finite dimension [12, 25, 31]. Then $\psi_k \rightarrow 0$ holds - by definition - for all choices of v_k in (3.11), i.e., for all $v_k = f(x_{k-1}) - f(x_k) + A(x_k - x_{k-1})$ with $A \in \partial f(x_{k-1})$.

In the same way, particular Newton maps $G(x) = \mathcal{N}f(x)$ can be handled [24].

However, for obtaining *Kantorovich-type convergence* if $f \notin C^1$, additional estimates and hypotheses will be needed for deriving the above C^1 -facts, and the interrelation between $\|Df(x_0)^{-1}\|$, δ and Λ must be appropriately specified.

- For *single-valued mappings* $G(x)$, one of these hypotheses is the approximation condition

$$\|f(x) - G(\xi)(x - \xi) - (f(x') - G(\xi)(x' - \xi))\| \leq \delta(r) \|x - x'\| \quad (3.26)$$

$$\forall x, x', \xi \in B(x_0, r) \subset \Omega \quad \text{where } \delta(r) \downarrow 0 \text{ as } r \downarrow 0,$$

which means equivalently with Remark 1 for the approximation functions

$$x \mapsto g_\xi(x) := \Sigma(x, \xi) - f(x) \quad (3.27)$$

and the same function $\delta = \delta(r)$, that

$$\|g_\xi(x) - g_\xi(x')\| \leq \delta(r) \|x - x'\| \quad \forall x, x', \xi \in B(x_0, r) \subset \Omega. \quad (3.28)$$

Setting $x' = \xi = x_{k-1} \in \Omega$ in (3.26), now the mean-value relation (3.13) follows globally on $B(x_0, r) \subset \Omega$,

$$\|f(x) - f(x_{k-1}) - G(x_{k-1})(x - x_{k-1})\| \leq \delta(r)\|x - x_{k-1}\| \quad (3.29)$$

$$\forall x, x_{k-1} \in B(x_0, r) \subset \Omega \text{ where } \delta(r) \downarrow 0 \text{ as } r \downarrow 0,$$

and the functions (3.27) are uniformly Lipschitz with small rank $\delta(r)$. In consequence, the conditions (3.24) and (3.26) ensure once more the uniform Lipschitz estimate (or contractivity) (3.15) for the quantities in (3.10). Hence the related convergence statements are valid whenever r is small enough.

- For multi-valued $G(x)$, the analogue condition to (3.26) becomes

$$d_H(\Sigma(x, \xi) - f(x), \Sigma(x', \xi) - f(x')) \leq \delta(r)\|x - x'\| \quad \forall x, x', \xi \in B(x_0, r) \subset \Omega \quad (3.30)$$

with the Hausdorff distance d_H , and non-empty images of $G(x)^{-1}$ must be assumed.

In any case, it remains a key problem to find concrete pairs (f, G) which satisfy the crucial conditions (3.26) or (3.30) and to understand the resulting auxiliary problems (1.7).

Next we show that the equivalent conditions (3.26) and (3.28) occur in a canonical manner as (3.40) and (3.39) below in the context of point-based approximations.

3.3 Point based approximations and C^1 points

When $f \notin C^1$ and G replaces Df , one needs analogons in terms of G for the estimate of $\|Df(x) - Df(x_0)\|$ and of $\|Df(x)^{-1}\| - \|Df(x_0)^{-1}\|$ as in Lemma 3. The constant L becomes a (luLip) rank of $G(x_0)^{-1}$.

What can be done for this aim is already visible by S.M. Robinson's work [35, 37]. First collect a set of axioms for f and G - there in form of a point-based approximation (PBA) - which ensures the convergence in question. As a second step (we call it justification), present a concrete class of functions and "derivatives" G which can be handled by this approach, cf. (3.35), and explain the related auxiliary problems (1.7). For the composed functions satisfying (3.35) below, provided they are not a-priori of type C^1 , the latter are always *piecewise linear* systems, cf. (3.36).

Such systems are also crucial for A.O. Griewank's approach [14–16]: Assume f is composed by differentiable standard functions and simple non-differentiable functions like *abs* or *max*. Apply an automatic differentiation program to f where *abs* or *max* remain unchanged. The result is (in the best case) a piecewise linear function.

Thus, for both concrete situations, the auxiliary problems are piecewise linear equations, even if the current iteration x_k is a C^1 -point of f .

Following Robinson (with the notation $\Sigma(x', x) = A(x, x')$ in [35, 37]) a function $\Sigma : \Omega \times \Omega \rightarrow Y$ is called a *point based approximation (PBA)* for f on $\Omega \ni x_0$, if there is a constant K such that, for all $x, \xi, x', x'' \in \Omega$,

$$\begin{aligned} (a) \quad & \|f(x') - \Sigma(x', x)\| \leq \frac{1}{2}K \|x' - x\|^2 \\ (b) \quad & \|\Sigma(x'', x) - \Sigma(x'', \xi) - [\Sigma(x', x) - \Sigma(x', \xi)]\| \leq K \|x - \xi\| \|x'' - x'\|. \end{aligned} \quad (3.31)$$

Using Remark 1 and

$$g_x(x') = f(x') - \Sigma(x', x) = f(x') - f(x) - G(x)(x' - x) \quad (3.32)$$

condition (3.31) means in terms of G

$$\begin{aligned} (a') \quad & \|g_x(x')\| \leq \frac{1}{2}K \|x' - x\|^2 \\ (b') \quad & \|G(x)(x'' - x) - G(x)(x' - x) + G(\xi)(x' - \xi) - G(\xi)(x'' - \xi)\| \leq K \|x - \xi\| \|x'' - x'\|. \end{aligned}$$

If Df exists at x and ξ and $G = Df$ holds at both points, condition (b') turns into

$$\|Df(x) - Df(\xi)\| \leq \delta_{x, \xi} := K\|x - \xi\| \quad (3.33)$$

and specifies condition (3.16) of Prop. 1.

Robinson's original result [37, Thm. 3.2] supposes further continuity of f and, instead of $\|Df(x_0)^{-1}\| \leq L$ in Prop. 1, strong regularity (cf. §2) of the function $x \mapsto \Sigma(x, x_0)$ at $(x_0, 0)$ with rank L ($= d_0^{-1}$ in condition b of Thm. 3.2). The latter coincides by (a) with strong regularity of f at $(x_0, f(x_0))$, and his crucial condition becomes, like (3.23),

$$L K \|x_1 - x_0\| = L \delta_{x_1, x_0} < \frac{1}{2} \quad (3.34)$$

for showing that procedure (1.4) ensures superlinear convergence to a zero $x^* \in \Omega$.

For the proof, the exact interplay of his constants, the needed size of $\Omega \ni x_0$ and estimates of the form $\|x_k - x^*\| \leq c \theta^{(2^k)} / (1 - \theta^{(2^k)})$; $\theta < 1$, we refer to the skillful original papers [35, 37] (which essentially fall back on a proof technique from [30]).

Additionally, it is worth to mention that Lemma 3.1 in [36] (cited in [37, Lemma 2.3]) proves persistence of strong regularity with respect to small Lipschitzian perturbations of continuous functions; a fact which has been generalized in terms of several *perturbation theorems* (extending Banach's Perturbation Lemma) for multifunctions later. Many facts concerning such perturbations in view of strong or metric regularity can be found in [17, §3.3]; we refer exemplarily also to [9, Lemma 3], [10, §§3F–3.G], [11, Thm. 4.1], [18, chap. 2] or [23, chap. 4].

PBAs for $f = h \circ \gamma$. Without having a concrete PBA, one cannot formulate or solve the auxiliary problems. So it was shown in [35, §2] that (3.31) is satisfied for composed functions

$$\begin{aligned} f &= h \circ \gamma \quad \text{where} \quad \gamma : \Omega \rightarrow Z, \quad h : Z \rightarrow Y, \\ \text{Lip}(\gamma, \Omega) &\leq L_\gamma \quad \text{and} \quad h \text{ has derivatives } Dh \text{ on an open set} \\ &\text{containing } \gamma(\Omega) \text{ which are Lipschitz with modulus } L_{Dh}. \end{aligned} \quad (3.35)$$

The proof uses the mean-value theorem with respect to h and keeps γ unchanged, and it clarifies that $K = L_{Dh} L_\gamma^2$ fulfills (3.31) for $\Sigma(x', x) = f(x) + Dh(\gamma(x))(\gamma(x') - \gamma(x))$. Hence Robinson's method for such f means to solve, with $D^0\gamma$ according to (1.8),

$$0 = f(x_k) + Dh(\gamma(x_k))(\gamma(x_{k+1}) - \gamma(x_k)) = f(x_k) + Dh(\gamma(x_k))(D^0\gamma(x_k)(x_{k+1} - x_k)). \quad (3.36)$$

In that context, projections γ onto closed convex sets in Hilbert spaces can be handled. They allow that $f = 0$ describes variational conditions. If γ is piecewise linear, (3.36) is always a piecewise linear equation for x_{k+1} . Unfortunately, one cannot solve *any Lipschitz equation* $\gamma = 0$ by the device of setting $h(x) = x$ in (3.35), since (3.36) leads us again to $\gamma(x_{k+1}) = 0$. Supposing the particular form (3.35) of f , a Kantorovich-type statement, also for generalized equations $0 \in f + M$ and assuming only $h \in C^1$, has been shown and discussed in [24].

Handling C^1 -points. Next suppose the general axioms (3.31) of a PBA and require in addition

$$\text{Assumption (A1): } \Sigma(x', x) = f(x) + Df(x)(x' - x) \quad \text{if } x \text{ is a } C^1\text{-point of } f.$$

Then, if x is a C^1 -point of f , condition (b') becomes

$$\|Df(x)(x'' - x') + G(\xi)(x' - \xi) - G(\xi)(x'' - \xi)\| \leq K \|x - \xi\| \|x'' - x'\|$$

for all $\xi, x', x'' \in \Omega$ and (a') requires in particular

$$\|g_\xi(x)\| \leq \frac{1}{2} K \|\xi - x\|^2 \quad \forall \xi \in B(x_0, r).$$

Furthermore, ξ may be another C^1 point. Then (b') and (a') turn into

$$\|Df(x) - Df(\xi)\| \leq K \|x - \xi\| \quad \text{and} \quad \|g_\xi(x)\| = \|f(x) - f(\xi) - Df(\xi)(x - \xi)\| \leq \frac{1}{2} K \|\xi - x\|^2.$$

Taking $x = \varepsilon$ and $\xi = -\varepsilon$ for $f = |\cdot|$, one sees that both conditions will generally fail to hold for piecewise linear or PC^1 functions, satisfying (3.35) or not. Hence, for such functions, there is no serious hope to find feasible definitions of PBAs which satisfy Assumption (A1).

Weaker PBAs. In consequence, let us weaken the PBA-properties by requiring weaker approximations in (a) and (b) which are still satisfied for the functions $f = h \circ \gamma$ of (3.35).

(P1) For the function g_ξ in (3.32), it follows obviously, for small r with $B(x_0, r) \subset \Omega$, that

$$\|g_\xi(x)\| = \|f(x) - f(\xi) - G(\xi)(x - \xi)\| \leq \delta(r)\|x - \xi\| \quad \forall x, \xi \in B(x_0, r) \quad (3.37)$$

holds with a function δ such that $\delta(r) \downarrow 0$ as $r \downarrow 0$ (e.g., $\delta(r) = \frac{1}{2}Kr$) and $g_\xi(\xi) = 0$.

(P2) By mean-value estimates as in [24, §2.3], the function g_ξ for (3.35) fulfills also the *uniform Lipschitz property* (3.28), i.e.,

$$\|g_\xi(x') - g_\xi(x'')\| \leq \delta(r)\|x' - x''\| \quad \forall x', x'', \xi \in B(x_0, r) \subset \Omega. \quad (3.38)$$

Setting here $x'' = \xi$, one sees that (3.37) follows from (3.38), so the latter is the crucial condition. Using the identity (3.32), inequality (3.38) also means (with $x'' = x$)

$$\|(\Sigma(x, \xi) - f(x)) - (\Sigma(x', \xi) - f(x'))\| \leq \delta(r)\|x - x'\| \quad \forall x, x', \xi \in B(x_0, r) \subset \Omega, \quad (3.39)$$

and coincides with the already mentioned general approximation condition (3.26),

$$\|f(x) - G(\xi)(x - \xi) - (f(x') - G(\xi)(x' - \xi))\| \leq \delta(r)\|x - x'\| \quad \forall x, x', \xi \in B(x_0, r) \subset \Omega. \quad (3.40)$$

The equivalent conditions (3.39) and (3.40) are obviously weaker than (a) and (b) in (3.31) and connect again the properties of f , G and Σ for small $\delta(r)$.

They appear as main assumptions for Kantorovich-type convergence in many subsequent papers, e.g. in [5, Thm. 2.2 and Corollary 2.5] with any function Σ or with directional derivatives $G(x)$ in [26, Thm. 4 (iii)].

The crucial condition (3.39) is also supposed in [1, Thm. 6.3 (ii)] for *multivalued* Σ where (3.39) attains the form (3.30). In addition, the main hypotheses

$$L\delta(r) < 1 \quad \text{and} \quad L\delta(r)\|x_1 - x_0\| < 1 \quad (3.41)$$

respectively, are required for the convergence theorems with some (psLip) rank L .

But disregarding the concrete situation (3.35), the *meaning* of the assigned auxiliary problems (1.7) - based on arbitrary PBAs or on methods which suppose directly the necessary consequences (3.39) of particular PBAs - is nowhere discussed; at least in all references we found. We also did not find any concrete PBA Σ if f differs from $h \circ \gamma$.

3.4 Sufficient conditions by supposing metric regularity

For extending Prop. 1 to model (1.7) it suffices again that all (or some appropriate selections of) iterations $x_{k+1} \in \mathcal{S}(x_k)$ converge to a *fixed point* x^* of \mathcal{S} . This can be ensured by the (psLip) property of \mathcal{S} with rank < 1 and holds generally, like Lemma 1, for any mapping $S : X \rightrightarrows X$. One has only to ensure that the points cannot leave the region of (psLip), cf. Def. 1.

Proposition 2 (*Big ball Prop.*) *Let $S : X \rightrightarrows X$ be arbitrary and suppose that S is (psLip) at $(x_0, x_1) \in \text{gph } S$ with rank $q < 1$ and assigned balls $B(x_1, r_1), B(x_0, r_2)$ satisfying the big ball condition*

$$\frac{\|x_1 - x_0\|}{1 - q} \leq r_2 \quad \text{and} \quad \frac{2 - q}{1 - q} \|x_1 - x_0\| \leq r_1. \quad (3.42)$$

Then there are $x_{k+1} \in S(x_k)$ such that $x_k \rightarrow x^$ in $\Omega_q := B(x_0, \frac{\|x_1 - x_0\|}{1 - q})$, and the convergence condition (3.1) holds true. If $S(x^*)$ is closed then $x^* \in S(x^*)$ and, if S is the solution map \mathcal{S} in (2.2), also $f(x^*) = 0$ follows.*

Note. The proposition concerns also zeros of particular sums of multifunctions due to

$$\begin{aligned} 0 \in (S - id)(x^*) &\Leftrightarrow x^* \in S(x^*) \Leftrightarrow x^* \in S^{-1}(x^*) \\ &\Leftrightarrow 0 \in (S^{-1} - id)(x^*) \Leftrightarrow x^* \in (S^{-1} - id)^{-1}(0). \end{aligned}$$

Proof We apply Def. 1 to S at $(\bar{y}, \bar{x}) = (x_0, x_1)$ with the particular points $x = y'$ and $y = x'$:

$$\begin{aligned} &\text{Some } x'' \in S(x) \text{ fulfills } \|x'' - x\| \leq q \|x - x'\| \\ &\text{if } x \in S(x'), \ x \in B(x_1, r_1) \text{ and } x, x' \in B(x_0, r_2). \end{aligned} \quad (3.43)$$

Thus, if $\Omega_q \subset [B(x_1, r_1) \cap B(x_0, r_2)]$, condition (3.3) of Lemma 1 is fulfilled and yields convergence of some related sequence $\{x_k\}$. The inclusion holds for $d := \frac{\|x_1 - x_0\|}{1-q} \leq r_2$ and $B(x_0, d) \subset B(x_1, r_1)$, hence if $d \leq r_2$ and $\|x_1 - x_0\| + d \leq r_1$. Explicitly, this is just (3.42).

If $S(x^*)$ is closed, put $y' = x^*$, $y = x_k$, $x = x_{k+1}$ in Def. 1. Then some $x'_k \in S(x^*)$ fulfills $\|x'_k - x_{k+1}\| \leq q \|x^* - x_k\|$. Thus the limit $x^* = \lim x_{k+1} = \lim x'_k$ belongs to the closed set $S(x^*)$. The latter yields $f(x^*) = 0$ for $S = \mathcal{S}$, so the proof is complete.

The requirement $q < 1$ is crucial for the proposition and must be ensured by further appropriate assumptions.

With the composed mapping in (2.2), $\mathcal{S}(x) = x + G(x)^{-1}(-f(x))$, the related Lipschitz property with rank $q_0 < 1$ for

$$\mathcal{S}_0(x) = x + G(x_0)^{-1}(-f(x)),$$

and a certain *regularity* can be supposed. Provided that some appropriate “difference” between $G(x), G(x_0)$ is small enough (comparable with δ in Prop. 1) then Prop. 2 still holds with some $q \in (q_0, 1)$. Necessary estimates have been presented in the framework of variational analysis - mostly in terms of metric regularity of $\mathcal{S}(x)^{-1}$ - by perturbation theorems where the explicit *stability balls* are often hidden in the proofs, cf. e.g. [10, 17, 18, 23].

In any case, by using appropriate (uniform) regularities for $G(x)$, $G(x)^{-1}$ and f , like (2.3), (2.4), so *sufficient conditions* for the desired convergence to a solution in $\Omega = \Omega_q$ can be described. This is also the key of recent generalizations [1, 5, 8] which derive Kantorovich-type statements for nonsmooth functions f and/or generalized equations (4.6) with closed mappings M .

4 Necessary conditions and problems of the approaches

4.1 Linear Newton-auxiliary problems for nonsmooth f

Let us now turn to the question whether also linear auxiliary problems can be used in order to establish Kantorovich-type statements for general methods in nonsmooth models. To find an answer, let us again suppose Assumption (A1): Apply usual Newton steps at C^1 -points $\xi = x_k \in \Omega$ of f .

For such $\xi \in B(x_0, r)$, we have $\Sigma(x', \xi) = f(\xi) + Df(\xi)(x' - \xi)$ and (3.39) requires

$$\|f(x) - f(x') - Df(\xi)(x - x')\| \leq \delta(r) \|x - x'\| \quad \forall x, x' \in B(x_0, r) \subset \Omega \quad (4.1)$$

$$\text{i.e.,} \quad \left\| \frac{f(x) - f(x')}{\|x - x'\|} - Df(\xi) \frac{x - x'}{\|x - x'\|} \right\| \leq \delta(r) \quad \forall x, x' \in B(x_0, r) \subset \Omega, \ x \neq x'. \quad (4.2)$$

Hence for all C^1 -points $\xi \in B(x_0, r)$ and related $x' \neq x$, the *difference quotients* $\frac{f(x) - f(x')}{\|x - x'\|}$ may differ from the *derivative-based approximation* $Df(\xi) \frac{x - x'}{\|x - x'\|}$ by $\delta(r)$ only.

Due to the additionally needed assumption $\delta(r) < L^{-1}$, see (3.41), this requires that the “size of all kinks” of f on $B(x_0, r)$ is small enough.

But it is just a typical property of piecewise differentiable functions that $\delta(r) \downarrow 0$ need not to hold (as in our Example 1 below, where (4.2) turns into $|3 - 1| \leq \delta(r)$). The same unpleasant situation occurs under Assumption (A1) for *any locally Lipschitz function f with a dense set of C^1 points*. This shrinks not only possible applications. It induces also the question of what else the methods should do at C^1 points x_k . In other words, Σ must be specified in an appropriate manner, in order to ensure that (3.38) or at least $\delta(r) < L^{-1}$ (not to speak about other assumptions) is a realistic hypothesis.

Under Assumption (A1), a Kantorovich-type statement (or local convergence at all) need not be true. This happens even for *real locally Lipschitz functions* and with the adapted requirement that, instead of δ , only the product $\delta \|x_1 - x_0\|$ is sufficiently small, like under (3.17) or (3.23).

Under [23, Example BE.1] one finds a real Lipschitz function (not semismooth with globally Lipschitzian inverse f^{-1}) such that the (usual) Newton methods generates alternating sequences $x_0, x_1, x_2 = -x_1, x_3 = x_1, \dots$ whenever the initial point $x_0 \neq x^*$ is arbitrarily chosen in the (open and dense) set of C^1 points. There, also x_1 and x_2 are C^1 points and $|x_1 - x_0| \rightarrow 0$ as $x_0 \rightarrow x^*$ holds true.

Here, we study simpler functions which are semismooth, depend on some parameter α and are piecewise linear near the unique zero $x^* = 0$.

Example 1 Given any $\alpha > 0$ let I_1, I_2 be the intervals $I_1 = [0, \alpha]$, $I_2 = [\alpha, 2\alpha]$ and define a function s on $I = I_1 \cup I_2$ by

$$s(x) = \begin{cases} x & \text{if } x \in I_1 \\ 2\alpha - x & \text{if } x \in I_2 \end{cases}.$$

On all intervals $I^p = r_p + I$ shifted by $r_p = 2p\alpha$ (p integer), we define a zigzag-function s periodically

$$s(x + r_p) = s(x) \quad \forall x \in I \quad \forall p$$

and the strongly regular function

$$f(x) = 2x + s(x)$$

with $\text{Lip}(f^{-1}, \mathbb{R}) = 1$ and the unique zero $x^* = 0$. Obviously, f and s have kinks at r_p and at all center points $c_p = 2p\alpha + \alpha$ of I^p , and it holds $s(c_p) = \alpha$, $s(r_p) = 0$ as well as

$$\begin{aligned} x \in \text{int } I_1^p &\Rightarrow s(x) = x - 2p\alpha, & f(x) = 3x - 2p\alpha, & Df = 3, \\ x \in \text{int } I_2^p &\Rightarrow s(x) = \alpha + c_p - x, & f(x) = x + \alpha + c_p, & Df = 1. \end{aligned} \quad (4.3)$$

Thus $Df(x) \in \{1, 3\}$ holds at all C^1 -points x and Df jumps by $\delta_0 = 2$ at each kink $x = c_p$ and $x = r_p$. There, also $\partial^{C^1} f(x) = [1, 3]$ holds true.

The parameter $\alpha > 0$ describes the *number* of kinks in any given interval Ω , their size $\delta_0 = 2$ remains fixed for all α .

Let us discuss Newton's method for this example. Under (A1), the iterates at C^1 -points x_k are $x_{k+1} = x_k - Df(x_k)^{-1} f(x_k)$ and (4.3) yields explicitly

$$\begin{aligned} x_k \in \text{int } I_1^p &\Rightarrow Df = 3, x_{k+1} = x_k - \frac{1}{3}(3x_k - 2p\alpha) = \frac{2}{3}p\alpha \\ x_k \in \text{int } I_2^p &\Rightarrow Df = 1, x_{k+1} = x_k - (x_k + \alpha + c_p) = -(\alpha + c_p) = -(2p+2)\alpha. \end{aligned} \quad (4.4)$$

- If one applies the semismooth Newton method and selects the Clarke-Jacobian 1 at all r_p , one obtains, beginning with any $x_0 = r_p$, the alternating sequence $x_0, -x_0, x_0, -x_0, \dots$ due to $f(r_p) = 2r_p$ and $x_1 = r_p - 2r_p$. Hence convergence then requires $x_0 = r_p = 0$. This shows that the semismooth method converges only for $|x_0 - x^*| < 2\alpha$ (if at all).

- If, with small $\varepsilon > 0$ and $p \geq 0$, the point $x_0 := 2(p+1)\alpha - \varepsilon$ is an initial point, then $x_0 \in \text{int } I_2^p$ and $x_1 = -2(p+1)\alpha$. Thus the condition for linear convergence is violated:

$$|x_1 - x^*|/|x_0 - x^*| = (2(p+1)\alpha)/(2(p+1)\alpha - \varepsilon) \geq 1.$$

The latter is true for the modified Newton method as well.

- Nevertheless, superlinear convergence with all initial points $x_0 \in \Omega := (-\alpha, \alpha)$ is ensured by one step only since now $x_1 = 0 = x^*$ follows from (4.4). The same interval gives us the region of convergence for the usual local semismooth Newton method. Because of

$$\Omega = \{x \mid -\alpha < f(x) < 3\alpha\},$$

formally also a Kantorovich-type theorem holds true: *Superlinear convergence to the (unique) solution is satisfied if $|f(x_0)|$ or $|x_1 - x_0|$ are small enough, e.g. $< \alpha$.*

However, given any semismooth or (more special) any PC^1 -functions f , neither the parameter α , which stands for the *structure* of kinks, nor the *size* δ_0 of kinks in some set Ω can be generally controlled. So one can also not check if $\Omega \ni x_0$ fulfills the requirements of a Kantorovich-type theorem.

4.2 General comments on inclusions as auxiliary problems

Let us mention some further principal problems of the multivalued approach, i.e., if G or the solution mapping \mathcal{S} are not single-valued.

1. The related Kantorovich-type theorem can be used (like implicit function theorems) for verifying the existence of a solution x^* and estimating its distance to the initial point x_0 . But it does not present a *solution method* as long as - in case of multivalued $\mathcal{S}(x_k)$ - the selection of an appropriate x_{k+1} is not described.

2. Even if all sets $\mathcal{S}(x)$ of auxiliary solutions are singletons, the following problem concerning G is not addressed in the literature: Every multifunction is a union of functions. So instead of $G(x) : X \rightrightarrows Y$, one may equivalently consider a related set $\tilde{G}(x)$ of functions $\gamma : X \rightarrow Y$ such that $G(x)(u) = \{\gamma(u) \mid \gamma \in \tilde{G}(x)\}$. Taking this into account, the auxiliary problems (1.4), (1.7) mean:

$$\text{find a particular element } \tilde{\gamma} \in \tilde{G}(x_k) \text{ and solve } f(x_k) + \tilde{\gamma}(x_{k+1} - x_k) = 0.$$

Other elements $\gamma \in \tilde{G}(x_k)$ may define other solutions x_{k+1} or the equation could be unsolvable at all. Thus we must not only solve a possibly nonlinear equation. We have also to find the "right" function $\tilde{\gamma} \in \tilde{G}(x_k)$ which ensures the needed estimates. For instance, by applying generalized Jacobians for $X = Y = \mathbb{R}^n$ and $G(x) = \partial f(x)$ as after Remark 1, the selected matrix $\gamma := A \in \tilde{G}(x_k) = \partial f(x_k)$ should at least be a regular one.

To avoid this problem, one has to ensure that the existence of x_{k+1} and the needed convergence remains true, uniformly for each selection of $\tilde{\gamma} \in \tilde{G}(x_k)$. This shrinks again the class of possible pairs (f, G) .

As already mentioned in Remark 4, the hypotheses for *local convergence* of the "semismooth Newton method" overcome this drawback, since under non-degeneracy of $\partial f(\bar{x})$ one may choose any $A \in \partial f(x_k)$ if $\|x_0 - \bar{x}\|$ is small enough, after which all A^{-1} exist and remain uniformly bounded. But this conclusion is not true if only $\|f(x_0)\|$ is small. For other "derivatives" $G(x)$ the problem does not vanish since sufficiently small $\|x_0 - \bar{x}\|$ should not be supposed in view of Kantorovich-type statements.

For instance, let $G(x)(u)$ denote approximations of any (generalized) directional derivatives $G_d(x)(u)$ where, e.g., $G(x)(u) = G_d(x)(u) + B(0, \varepsilon(x, u))$, and small $\varepsilon(x, u)$ stands for an assigned exactness. Then (1.7) says that our solutions x_{k+1} , which possibly converge in some required manner, are given by

$$0 \in f(x_k) + G_d(x_k)(x_{k+1} - x_k) + w_k \quad (4.5)$$

for some *particular* $w_k \in B(0, \varepsilon(x_k, x_{k+1} - x_k))$. But, even if

$$0 \in f(x_k) + G_d(x_k)(x_{k+1} - x_k) + B(0, \varepsilon(x_k, x_{k+1} - x_k))$$

has a unique solution x_{k+1} , we do not know which $w_k \in B(0, \varepsilon(x_k, x_{k+1} - x_k))$ can be chosen in (4.5) for the approximation as long as the related convergence is not ensured for all w_k .

4.3 The analogue approach to generalized equations

Consider a generalized equation

$$0 \in f(x) + M(x) \quad (4.6)$$

with (closed) mappings $M : X \rightrightarrows Y$. In the simplest case $f \in C^1$, one mostly uses Newton steps as

$$0 \in \Sigma_M(x_{k+1}, x_k) := f(x_k) + Df(x_k)(x_{k+1} - x_k) + M(x_{k+1}), \quad x_0 \text{ given} \quad (4.7)$$

as in [21, 33] for particular normal maps M . In terms of G (again assigned to f and Σ), the auxiliary problems for (4.6) then become

$$0 \in f(x_k) + G(x_k)(x_{k+1} - x_k) + M(x_{k+1}), \quad (k \geq 0). \quad (4.8)$$

In order to see that the problems discussed in §4.2 will not disappear, put $f = 0$. This usually yields to the auxiliary problems $0 \in M(x_{k+1})$ which are again trivial.

On the other hand, one can use a derivative-like approximation \mathcal{D} of M at x_k in order to solve

$$0 \in M(x_k) + \mathcal{D}M(x_k)(x_{k+1} - x_k).$$

However, identifying M with a function f , the question of reasonable $\mathcal{D}f$ is the same as above.

Finally, in order to see that the Kantorovich-type statement generalizes local statements on Newton's method, the imposed hypotheses should hold (at least) if the distance of the initial point x_0 to a zero \bar{x} is already small enough. For generalized equations, this is not automatically true, see example 5 in [24].

Nevertheless, in various applications of *local* nonsmooth methods, e.g., for typical situations in optimal control and differential variational inequalities, the sets $G(x_k)$ and the solution sets of (4.8) are singletons and the problem (4.8) is well-handable. This is known from the rich literature on local convergence of nonsmooth Newton methods, see, e.g., the monographs [10, 12, 19, 20, 23, 40].

Acknowledgement. The authors are indebted to the referees for the very careful reading of the manuscript and several valuable remarks.

References

1. Adly, S., Cibulka, R., van Ngai, H.: *Newton's method for solving inclusion using set-valued approximations*. SIAM J. Optim. 25, 159–184 (2015)
2. Aragon Artacho, F.J., Belyakov, M., Dontchev, A.L., Lopez, M.: *Local convergence of quasi-Newton methods under metric regularity*. Comput. Optim. Appl. 58, 225–247 (2014)
3. Argyros, I.K.: *Convergence and Applications of Newton-type Iterations*. Springer, New York (2008)
4. Aubin, J.-P., Ekeland, I.: *Applied Nonlinear Analysis*. Wiley, New York (1984)
5. Cibulka, R., Dontchev, A.L., Preininger, J., Roubal, T., Veliov, V.: *Kantorovich-type theorems for generalized equations*. J. Convex Analysis 168, 673–716 (2018)
6. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. Wiley, New York (1983)
7. Cominetti, R.: *Metric regularity, tangent sets and second-order optimality conditions*. Applied Mathematics and Optimization, 21: 265–287 (1990)
8. Dontchev, A.L.: *Local convergence of the Newton method for generalized equation*. C. R. Acad. Sci. Paris, Ser. I 322, 327–331 (1996)
9. Dontchev, A.L.: *Local analysis of a Newton-type method based on partial linearization*. Lectures in Applied Mathematics 32, 295–306 (1996)
10. Dontchev, A.L., Rockafellar, R.T.: *Implicit Functions and Solution Mappings. A View from Variational Analysis*. Springer, Dordrecht, Heidelberg, London, New York (2009) (2nd edition 2014)
11. Dontchev, A.L., Rockafellar, R.T.: *Newton's method for generalized equations: a sequential implicit function theorem*. Math. Program. Ser. B 123, 139–159 (2010)
12. Facchinei, F., Pang, J.-S.: *Finite-Dimensional Variational Inequalities and Complementary Problems*, Vol II. Springer, New York (2003)
13. Fusek, P., Klatte, D., Kummer, B.: *Examples and counterexamples in Lipschitz analysis*. Control and Cybernetics, Vol 31 No 3, pp. 471–492 (2002)
14. Griewank, A.: *On stable piecewise linearization and generalized algorithmic differentiation*. Optim. Methods Softw. 28, 1139–1178 (2013)
15. Griewank, A., Bernt, J.-U., Radons, M., Streubel, T.: *Solving piecewise linear systems in abs-normal form*. Linear Algebra and its Applications 471, 500–530 (2015)
16. Griewank, A., Lehmann, L., Radons, M., Streubel, T.: *An open Newton method for piecewise smooth functions*. preprint (August 2018), <https://www.researchgate.net/publication/326764737>
17. Ioffe, A. D.: *Metric regularity and subdifferential calculus*. Uspekhi Mat. Nauk 55:3, 103–162 (2000)
18. Ioffe, A. D.: *Variational Analysis of Regular Mappings: Theory and Applications*. (Springer monographs in mathematics). Springer International Publisher, Cham (2017)
19. Ito, K., Kunisch, K.: *Lagrange Multiplier Approach to Variational Problems and Applications*. SIAM, Philadelphia (2008)
20. Izmailov, A.F., Solodov, M.V.: *Newton-Type Methods for Optimization and Variational Problems*. Springer, Cham (2014)
21. Josephy, N.H.: *Newton's method for generalized equations and the PIES energy model*. Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin-Madison (1979)
22. Kantorovich, L.W., Akilov, G.P.: *Funktionalanalysis in normierten Räumen*. Akademie Verlag, Berlin (1964)
23. Klatte, D., Kummer, B.: *Nonsmooth Equations in Optimization - Regularity, Calculus, Methods and Applications*. Kluwer Academic Publ., Dordrecht-Boston-London (2002)

24. Klatte, D., Kummer, B.: *Approximations and generalized Newton methods*. Math. Program. Ser. B, 168, 673–716 (2018)
25. Kummer, B. *Newton's method for non-differentiable functions*. In Guddat, J. et al. (eds.) *Advances in Mathematical Optimization*, Ser. Math. Res. 45, 114–125. Akademie Verlag, Berlin (1988)
26. Kummer, B. *On stability and Newton-type methods for Lipschitzian equations with applications to optimization problems*. In P. Kall ed., *System Modelling and Optimization*, Proceedings of the 15th IFIP Conference, Zurich 1991. Springer, 3–16 (1991)
27. Mordukhovich, B.S.: *Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions*. Trans. Amer. Math. Soc. 340, 1–35 (1993)
28. Mordukhovich, B.S. *Variational Analysis and Generalized Differentiation*, I, II. Springer, Berlin, Heidelberg, 2006.
29. Ortega, J.M.: *The Newton-Kantorovich theorem*. Amer. Math. Monthly 75, 658–660 (1968)
30. Potra, F.A., Ptak, V.: *Nondiscrete Induction and Iterative Processes*. Research Notes in Mathematics 103, Pitman, Boston MA (1984)
31. Qi, L., Sun, L.: *A nonsmooth version of Newton's method*. Math. Program. 58, 353–367 (1993)
32. Rheinboldt, W.C., *A unified convergence theorem for a class of iterative processes*. SIAM J. Numer. Anal. 5, 42–63 (1968)
33. Robinson, S.M.: *An implicit-function theorem for generalized variational inequalities*. Technical Summary Report No. 1672, Mathematics Research Center, University of Wisconsin-Madison (1976)
34. Robinson, S.M.: *Strongly regular generalized equations*. Math. Oper. Res. 5, 43–62 (1980)
35. Robinson, S.M.: *Newton's method for a class of nonsmooth functions*. Technical Report, Department of Industrial Engineering, University of Wisconsin-Madison (1988)
36. Robinson, S.M.: *An implicit-function theorem for a class of nonsmooth functions*. Math. Oper. Res. 16, 292–309 (1991)
37. Robinson, S.M.: *Newton's method for a class of nonsmooth functions*. Set-Valued Anal. 2, 291–305 (1994)
38. Rockafellar, R.T., Wets, J.-B.: *Variational Analysis*. Springer, Berlin (1998)
39. Scholtes, S.: *Introduction to Piecewise Differentiable Equations*. Springer, New York (2012) (Preprint 1998)
40. Ulbrich, M.: *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. SIAM, Philadelphia (2011)